Journal of Medical Systems (2025) 49:51
https://doi.org/10.1007/s10916-025-02177-0

RESEARCH

# Development of an Artificial Intelligence-Enabled Electrocardiography to Detect 23 Cardiac Arrhythmias and Predict Cardiovascular Outcomes

**Wen-Yu Lin[1] · Chin Lin[2,3,4] · Wen-Cheng Liu[1] · Wei-Ting Liu[1] · Chiao-Hsiang Chang[1] · Hung-Yi Chen[5] · Chiao-Chin Lee[1] · Yu-Cheng Chen[5] · Chen-Shu Wu[5] · Chia-Cheng Lee[6,7] · Chih-Hung Wang[8,9] · Chun-Cheng Liao[10,11,12,13] · Chin-Sheng Lin[1,14]**

Received: 25 February 2024 / Accepted: 22 March 2025


**Abstract**
Arrhythmias are common and can affect individuals with or without structural heart disease. Deep learning models (DLMs) have shown the ability to recognize arrhythmias using 12-lead electrocardiograms (ECGs). However, the limited types of arrhythmias and dataset robustness have hindered widespread adoption. This study aimed to develop a DLM capable of detecting various arrhythmias across diverse datasets. This algorithm development study utilized 22,130 ECGs, divided into development, tuning, validation, and competition sets. External validation was conducted on three open datasets (CODE-test, PTB-XL, CPSC2018) comprising 32,495 ECGs. The study also assessed the long-term risks of new-onset atrial fibrillation (AF), heart failure (HF), and mortality in individuals with false-positive AF detection by the DLM. In the validation set, the DLM achieved area under the receiver operating characteristic curve above 0.97 and sensitivity/specificity exceeding 90% across most arrhythmia classes. It demonstrated cardiologist-level performance, ranking first in balanced accuracy in a human-machine competition. External validation confirmed comparable performance. Individuals with false-positive AF detection had a significantly higher risk of new-onset AF (hazard ration [HR]: 1.69, 95% confidence interval [CI]: 1.11–2.59), HF (HR: 1.73, 95% CI: 1.20–2.51), and mortality (HR: 1.40, 95% CI: 1.02–1.92) compared to true-negative individuals after adjusting for age and sex. We developed an accurate DLM capable of detecting 23 cardiac arrhythmias across multiple datasets. This DLM serves as a valuable screening tool to aid physicians in identifying high-risk patients, with potential implications for early intervention and risk stratification.

**Keywords** Artificial intelligence · Electrocardiogram · Deep learning · Arrhythmias · Atrial fibrillation

## Introduction

Cardiac arrhythmias include all conditions in which the heart's conduction system is disrupted. Structural or electrical abnormalities in the cardiomyocytes can lead to abnormal impulse formation or alter cardiac propagation, facilitating arrhythmias [1]. Arrhythmias are prevalent across all age groups and may occur in both healthy and structurally abnormal hearts [2, 3]. Cardiac arrhythmias are a major cause of morbidity and mortality worldwide, and the current approach to reducing adverse consequences relies on early recognition and treatment [4]. The importance of prompt and accurate arrhythmia diagnosis using various electrical recording devices is clear.

The 12-lead electrocardiogram (ECG) is the most commonly used diagnostic tool to detect cardiac arrhythmias in clinical practice. It efficiently and rapidly records a 10-second heart rate and rhythm duration, providing information on cardiac electrical properties [5]. Precise interpretation of the 12-lead ECG is crucial for diagnosing cardiac

Chun-Cheng Liao and Chin-Sheng Lin contributed equally to this work.

Extended author information available on the last page of the article

Published online: 22 April 2025

arrhythmias. However, many residents and young attending physicians struggle to improve their ability to interpret ECGs accurately and often require assistance from expert cardiologists for a correct diagnosis. Computerized ECG interpretation by the device has been developed since the 1960s and has become routine today [6]. While computerized interpretations provide diagnostic assistance based on traditional rule-based criteria, they still exhibit significant errors when applied to complex arrhythmias [7]. Since expert cardiologists may not always be available, there is a need for a more comprehensive diagnostic algorithm to improve the accuracy of ECG analysis and better support clinicians in managing life-threatening arrhythmias.

Artificial intelligence (AI) and machine learning (ML) in medicine are currently areas of intense exploration, transforming clinical practice, particularly in ECG interpretation and prediction. Recently, widely available digital ECG data and the algorithmic paradigm of deep learning models (DLMs) have been applied to improve the accuracy and scalability of automated ECG analysis [8]. Previous studies have demonstrated that computer-assisted classification and diagnosis of cardiac arrhythmias have provided comparable diagnostic accuracy to that of a cardiologist [9–12]. However, previous DLMs investigated only parts of common arrhythmias, potentially missing certain uncommon but clinically significant arrhythmias. In our study, we established and validated a comprehensive DLM capable of detecting 23 clinically important cardiac arrhythmias. To ensure the generalizability of our DLM, we performed external validation using three worldwide open-access databases.

## Materials and Methods

### Study Design and Data Collection

The study was an algorithm development study primarily using the Tri-Service General Hospital, Taiwan database to establish a DLM and assess its performance for cardiac arrhythmia detection. This study was approved by the Institutional Review Board of the Tri-Service General Hospital, National Defense Medical Center (IRB NO. C202105049), in accordance with the ethical guidelines of the Declaration of Helsinki of the World Medical Association. Patient consent was waived because data were collected retrospectively and stored in anonymized, encrypted files that were transferred from the hospital to the data controller. Research groups had no access to any patient identities. To maximize the inclusion of rare yet clinically significant arrhythmias in our dataset, 7,663 patients who had undergone electrophysiological study and catheter ablation or possessed a corresponding International Classification of Diseases Tenth Revision (ICD-10) code were included. Accordingly, 22,130 ECGs were used to develop the DLM and were divided into four private datasets (development set, tuning set, validation set, and competition set), as shown in Fig. 1. Most of the patient data used for model development involved Asian participants, 57.1% of whom were male, with a mean age of approximately 70 years. All ECG recordings were collected using a Philips 12-lead ECG machine (PH080A) with a sampling frequency of 500 Hz and 10 s recorded in each lead. Patient characteristics of sex and age were acquired from the electronic medical record.

### Classification of Cardiac Arrhythmias

This study included 23 arrhythmia classes, encompassing sinus rhythm (SR), sinus arrhythmia (SA), sinus bradycardia (SB), sinus tachycardia (ST), ectopic atrial rhythm (EAT), first-degree atrioventricular block (1AVB), 2:1 atrioventricular block (2:1AVB), second-degree atrioventricular block, Mobitz type I (2AVB1), second-degree atrioventricular block, Mobitz type II (2AVB2), complete AV block (CAVB), junctional rhythm/junctional bradycardia (JR), atrial fibrillation (AF), typical/atypical atrial flutter (AFL), paroxysmal supraventricular tachycardia (PSVT), Wolff-Parkinson-White syndrome with pre-excitation (WPW), nonspecific intraventricular conduction disturbance (NSIVCD), left fascicular block (LFB) including left anterior fascicular block (LAFB) and left posterior fascicular block (LPFB), complete right bundle branch block (CRBBB), complete left bundle branch block (CLBBB), pacemaker rhythm (PMR), premature atrial complex (PAC), premature ventricular complex (PVC), idioventricular rhythm or accelerated idioventricular rhythm (IVR), ventricular tachycardia/ventricular fibrillation (VT/VF), and noise signals. The diagnostic terms and definitions for each arrhythmia class were established based on previous recommendations for the standardization and interpretation of ECGs [13, 14].

### Protocol of ECG Annotation and Development of Deep Learning Model

In this study, five cardiologists served as annotators. Each ECG was annotated by three independent annotators who were unaware of any patient information. A committee convened to reach a final determination in cases of disparate results among the three annotations. Stringent criteria for consistent annotation were applied to ensure complete uniformity in the 23 arrhythmia classifications within multiple-choice questions. The committee included all five annotators and three additional experienced cardiologists, and the final annotations were derived following a comprehensive review of medical records and consensus among the

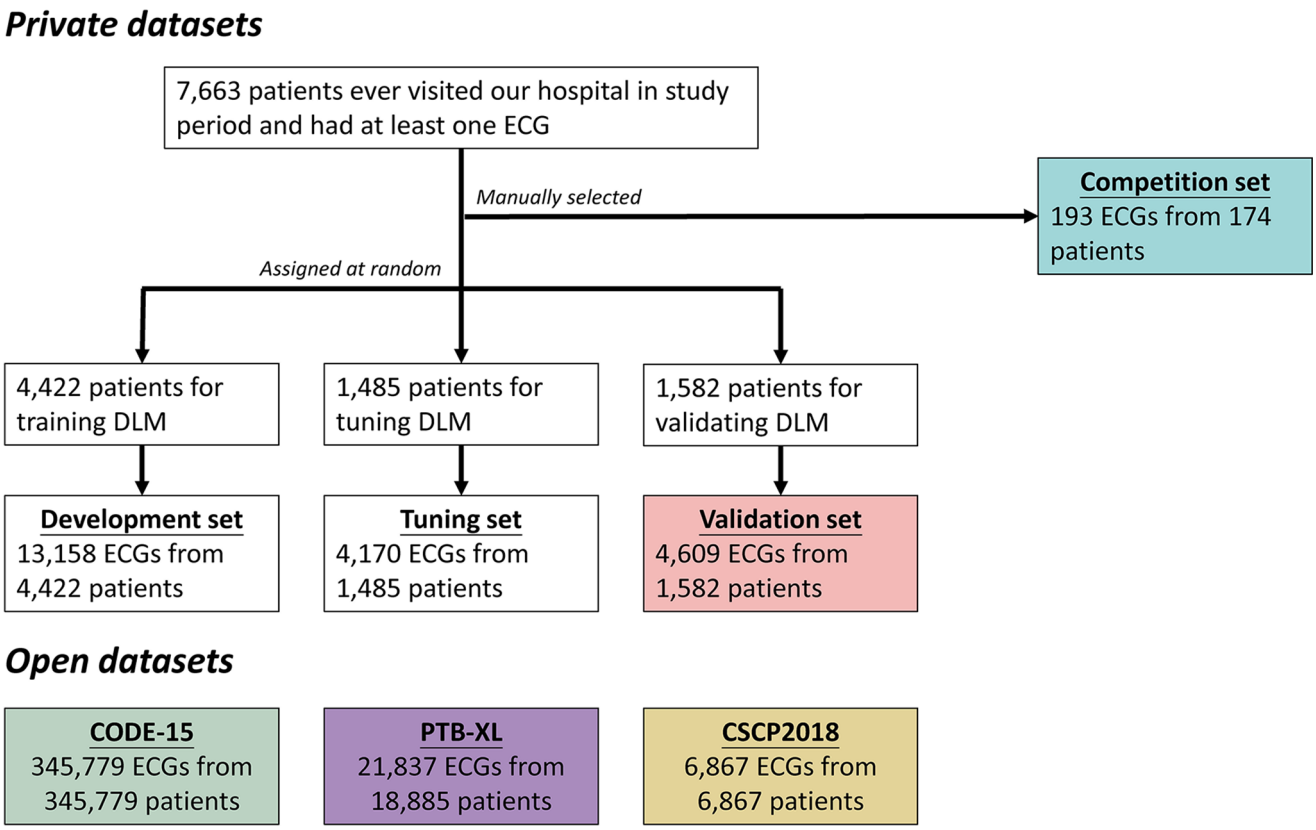


**Fig. 1** Development, tuning, internal validation, and external validation sets generation and ECG labeling of survival information in private dataset. Schematic of the dataset creation and analysis strategy, devised to ensure a robust and reliable dataset for model development, tuning, and validation of the network. Once a patient's data was placed in one of the private datasets, that individual's data was used only in that set, avoiding 'cross-contamination' among the model development, validation, and test datasets. The details of the flowchart and how each of the datasets was used are described in the Methods

cardiologists. The committee examined approximately 20% of ECGs with inconsistent annotations, and 193 ECGs from 174 patients were selected as an independent competition set for human-machine competition. To rigorously evaluate model performance across various clinical scenarios, the ECG selection process for human-machine competition prioritized arrhythmia diversity, patient representation, ECG quality, and a balance of easy and challenging cases. ECGs from the remaining patients were randomly assigned to a development set, a tuning set, and a validation set. The DLM underwent training and internal testing using the development set, and the tuning set was used to guide the training process. For each arrhythmia class, the validation set was employed for accuracy testing, and there were no overlaps between the private datasets.

To validate the DLM performance in external populations, we utilized three additional open datasets. The CODE test is a subset of the openly available Clinical Outcomes in Digital Electrocardiography (CODE) dataset (https://zenodo.org/record/3765780) [15]. The PTB-XL is a large publicly available electrocardiography multi-label dataset annotated by up to two cardiologists (https://physionet.org/content/ptb-xl/1.0.1/) [16]. The China Physiological Signal Challenge in 2018 (CPSC2018) (http://2018.icbeb.org/), which is openly available and was held during the 7th International Conference on Biomedical Engineering and Biotechnology in Nanjing, China [17]. We included only ECGs with signals lasting more than 10 s, resizing them to a sampling frequency of 500 Hz, consistent with our approach to processing our private datasets.

## Implementation of Deep Learning Model

To further enhance prediction performance, we incorporated ECG12Net, a 12-channel sequence-to-sequence model modified from DenseNet [18]. ECG12Net consists of 12 ECG lead blocks, each with 80 trainable layers that extract 864 features per lead to generate independent predictions. A hierarchical attention mechanism integrates information from all leads, improving interpretability and optimizing overall prediction accuracy [19].

We trained 25 independent DLMs for sinus rhythm, noise, and each arrhythmia class, and the architecture of the DLM was developed in our previous study. [20, 21]. Each

DLM produced a probability output describing a binary outcome, and we followed the same technological details as in training a new DLM [19, 22, 23]. The input format of the DLM was a sequence of 4,096 numeric values, but the original length of our 12-lead ECG signal was 5,000. During the training process, we randomly cropped a sequence of 4,096 as input. For the inference stage, two overlapping sequences of 4,096 at the start and end were used to generate predictions, which were averaged as the final prediction [24]. An oversampling process was implemented to ensure that rare arrhythmia classes were adequately recognized, sampling the same number of positive and negative ECGs in a single batch. We trained these DLMs with a batch size of 32 and used an initial learning rate of 0.001 with the Adam optimizer, using standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$). The learning rate was decayed by a factor of 10 each time the loss on the tuning set plateaued after an epoch. To prevent the DLM from overfitting, early stopping was performed by saving the network after every epoch and choosing the saved DLMs with the lowest loss on the tuning set. The only regularization method for avoiding overfitting was L2 regularization with a coefficient of $10^{-4}$ in this study.

### Outcome Variables

We conducted a follow-up on three AF-related outcomes of interest, including new-onset AF, heart failure (HF), and all-cause mortality, to explore the concept of "previvor" in AF within the validation set [25]. Patients without a history of AF at the time of initial classification were included in the analysis. New-onset AF was defined by the ICD-10 coded by physicians, and the new-onset HF was defined based on a quantitative ejection fraction of ≤35% routinely acquired by experienced cardiologists. For all-cause mortality, patient status (dead/alive) was extracted from the electronic medical record. Additionally, data for living patients were censored at the patient's last known hospital encounter to mitigate potential bias from incomplete records. The end of follow-up for all outcomes was September 30, 2021.

### Statistical Analysis

All analyses were based on ECGs. We presented their characteristics as means and standard deviations, numbers of patients, or percentages, where appropriate. The statistical analysis was carried out using the software environment R version 3.4.4. The software package MXNet version 1.3.0 was implemented in our deep learning model. We used a significance level of $p < 0.05$ throughout the analysis. The receiver operating characteristic (ROC) curve and area under the curve (AUC) were applied to evaluate the DLM performance. Additionally, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated based on the maximum of Youden's index in the tuning set, and all operating points were consistent in each dataset. In the human-machine competition, we included two visiting staff (one emergency physician and one cardiologist), three internal-medicine residents, and three post-graduate year (PGY) trainees. The physicians had no access to patient information for further diagnosis, and their sensitivity and specificity were calculated to compare with those of the DLM. The box plot and jittered points were used to present the distribution of probabilities given by DLMs in each annotated certainty group. For the "previvor" analysis, we performed Kaplan–Meier survival analysis with the available follow-up data stratified by the DLM prediction on each outcome of interest. Further, we fitted a Cox proportional-hazard model adjusted by sex and age. The hazard ratios (HRs) with 95% confidence intervals (95% CIs) were reported.

## Results

The baseline characteristics of each dataset are presented in Table 1. The development and tuning sets comprised a total of 13,158 ECGs with 57.1% males, aged 70.7 ± 17.5 years, and 4,170 ECGs with 57.5% males, aged 70.5 ± 16.9 years, respectively. Neither of the two sets was included in the performance assessment. The validation set, which had complete labels for 23 arrhythmia classes, consisted of 4,609 ECGs with 60.4% males, aged 70.4 ± 17.3 years. In the competition set, only 19 of the 23 arrhythmia classes were represented, comprising 193 ECGs with 66.3% males, aged 70.8 ± 17.6 years. Arrhythmias of EAT, 2AVB2, LFB, and IVR were excluded because of the limited number of samples, and NSIVCD was excluded due to unclear definitions. Additionally, all selected ECGs in the competition set exhibited clear ECG signals without noise. The open datasets, including CODE-test, PTB-XL, and CPSC2018 dataset, consisted of 827, 21,837, and 6,867 ECGs, respectively, with annotations of 6, 12, and 6 arrhythmia classes.

Figure 2 shows the performance of the DLM on the validation set. High-performance measures were obtained across 15 arrhythmia classes, including SR, SB, ST, 1AVB, 2:1AVB, CAVB, AF, AFL, PSVT, WPW, CLBBB, CRBBB, PMR, IVR, and VT/VF, with AUCs above 0.97 and sensitivity/specificity indexes exceeding 90%. Notably, the DLM performed better than the Philips ECG automatic analysis system [26], except for SA and 2AVB1. However, the sensitivity of the DLM remained superior to that of the Philips ECG automatic analysis system in these two arrhythmia classes. Given that annotators contributed to the standard answers in these ECGs, the DLM exhibited comparatively

**Table 1** Baseline characteristics in each dataset

| | Private datasets | | | | Open datasets | | |
|---|---|---|---|---|---|---|---|
| | Development ($n$=13158) | Tuning ($n$=4170) | Validation ($n$=4609) | Competition ($n$=193) | CODE-test ($n$=827) | PTB-XL ($n$=21837) | CPSC2018 ($n$=6867) |
| **Demography** | | | | | | | |
| Male | 6896(57.1%) | 2398(57.5%) | 2782(60.4%) | 128(66.3%) | 321(38.8%) | 10,458(47.9%) | 3692(53.8%) |
| Age | 70.7±17.5 | 70.5±16.9 | 70.4±17.3 | 70.8±17.6 | 54.9±16.5 | 59.8±17.0 | 60.2±19.1 |
| **Arrhythmia** | | | | | | | |
| SR | 3656(27.8%) | 1186(28.4%) | 1277(27.7%) | 36(18.7%) | | | |
| SA | 117(0.9%) | 31(0.7%) | 39(0.8%) | 10(5.2%) | | | |
| SB | 803(6.1%) | 260(6.2%) | 379(8.2%) | 11(5.7%) | 16(1.9%) | | |
| ST | 2138(16.2%) | 681(16.3%) | 716(15.5%) | 20(10.4%) | 37(4.5%) | | |
| EAT | 64(0.5%) | 19(0.5%) | 25(0.5%) | | | | |
| 1AVB | 2015(15.3%) | 723(17.3%) | 831(18.0%) | 40(20.7%) | 28(3.4%) | 797(3.6%) | 721(10.5%) |
| 2:1AVB | 92(0.7%) | 27(0.6%) | 44(1.0%) | 4(2.1%) | | | |
| 2AVB1 | 22(0.2%) | 4(0.1%) | 12(0.3%) | 6(3.1%) | | | |
| 2AVB2 | 39(0.3%) | 24(0.6%) | 34(0.7%) | | | | |
| CAVB | 263(2.0%) | 106(2.5%) | 103(2.2%) | 10(5.2%) | | 16(0.1%) | |
| JR | 401(3.0%) | 131(3.1%) | 127(2.8%) | 11(5.7%) | | | |
| AF | 2472(18.8%) | 830(19.9%) | 869(18.9%) | 16(8.3%) | 13(1.6%) | 1514(6.9%) | 1220(17.8%) |
| AFL | 347(2.6%) | 85(2.0%) | 114(2.5%) | 12(6.2%) | | 73(0.3%) | |
| PSVT | 303(2.3%) | 82(2.0%) | 95(2.1%) | 10(5.2%) | | 24(0.1%) | |
| WPW | 154(1.2%) | 43(1.0%) | 48(1.0%) | 10(5.2%) | | 80(0.4%) | |
| NSIVCD | 560(4.3%) | 207(5.0%) | 162(3.5%) | | | 789(3.6%) | |
| LFB | 102(0.8%) | 46(1.1%) | 56(1.2%) | | | | |
| CLBBB | 277(2.1%) | 93(2.2%) | 58(1.3%) | 11(5.7%) | 30(3.6%) | 536(2.5%) | 235(3.4%) |
| CRBBB | 1667(12.7%) | 476(11.4%) | 552(12.0%) | 40(20.7%) | 34(4.1%) | 542(2.5%) | 1854(27.0%) |
| PMR | 2526(19.2%) | 754(18.1%) | 811(17.6%) | 14(7.3%) | | 296(1.4%) | |
| PAC | 410(3.1%) | 167(4.0%) | 176(3.8%) | 23(11.9%) | | 398(1.8%) | 614(8.9%) |
| PVC | 881(6.7%) | 284(6.8%) | 280(6.1%) | 17(8.8%) | | 1146(5.2%) | 699(10.2%) |
| IVR | 26(0.2%) | 6(0.1%) | 8(0.2%) | | | | |
| VT/VF | 143(1.1%) | 43(1.0%) | 54(1.2%) | 7(3.6%) | | | |
| Noise | 55(0.4%) | 19(0.5%) | 11(0.2%) | | | | |

Abbreviations SR, sinus rhythm; SA, sinus arrhythmia; SB, sinus bradycardia; ST, sinus tachycardia; EAT, ectopic atrial rhythm; 1AVB, first-degree atrioventricular block; 2:1AVB, 2:1 atrioventricular block; 2AVB1, second-degree atrioventricular block (mobitz type I); 2AVB2, second-degree atrioventricular block (mobitz type II); CAVB, complete AV block; JR, junctional escape rhythm/junctional bradycardia; AF, atrial fibrillation; AFL, atrial flutter (typical/atypical); PSVT, paroxysmal supraventricular tachycardia; WPW, Wolff-Parkinson-White syndrome with pre-excitation; NSIVCD, nonspecific intraventricular conduction delay; LFB, left fascicular block; CLBBB, complete right bundle branch block; CRBBB, complete left bundle branch block; PMR, pacemaker rhythm; PAC, premature atrial complexes (bigeminy/trigeminy/quadrigeminy); PVC, premature ventricular complexes (bigeminy/trigeminy/quadrigeminy); IVR, idiopathic ventricular rhythm or accelerated idiopathic ventricular rhythm; VT/VF, ventricular tachycardia or ventricular fibrillation

lower performance than annotators across most arrhythmia classes. Nevertheless, the DLM still performed similarly to annotators in 2:1AVB, AF, WPW, CRBBB, PMR, and VT/VF.

The results of the human-machine competition are presented in Fig. 3. There were 8 participants in the competition, including three internal medicine residents, three PGY trainees, and two visiting staff (one emergency physician and one cardiologist). Using the same specificity as each participating physician, our DLM demonstrated superior or equivalent sensitivity in 2:1AVB, AF, CRBBB, and PMR compared to all participants. In the case of SB, ST, CAVB, JR, PSVT, WPW, CLBBB, and VT/VF, only three or fewer participants performed better than the DLM. For detecting SR, 1AVB, AFL, PAC, and PVC, the DLM ranked from fourth to sixth place compared to physicians but exhibited comparatively lower performance in detecting SA and 2AVB1 compared to the average level of participants. Regarding balanced accuracy, indicated by the mean of sensitivity and specificity, the DLM performed better than five or more participants in most arrhythmia classes, except for PVC and 2AVB1. Notably, in this human-machine competition, the Philips ECG automatic analysis system surpassed the DLM only in detecting SA, 2AVB1, and AFL. According to the average balanced accuracy across 19 arrhythmia classes, our DLM ranked first (92.3%), followed by a third-year cardiologist (86.0%), a fifteen-year emergency physician (84.7%), a second-year PGY-1 (83.7%), first-year

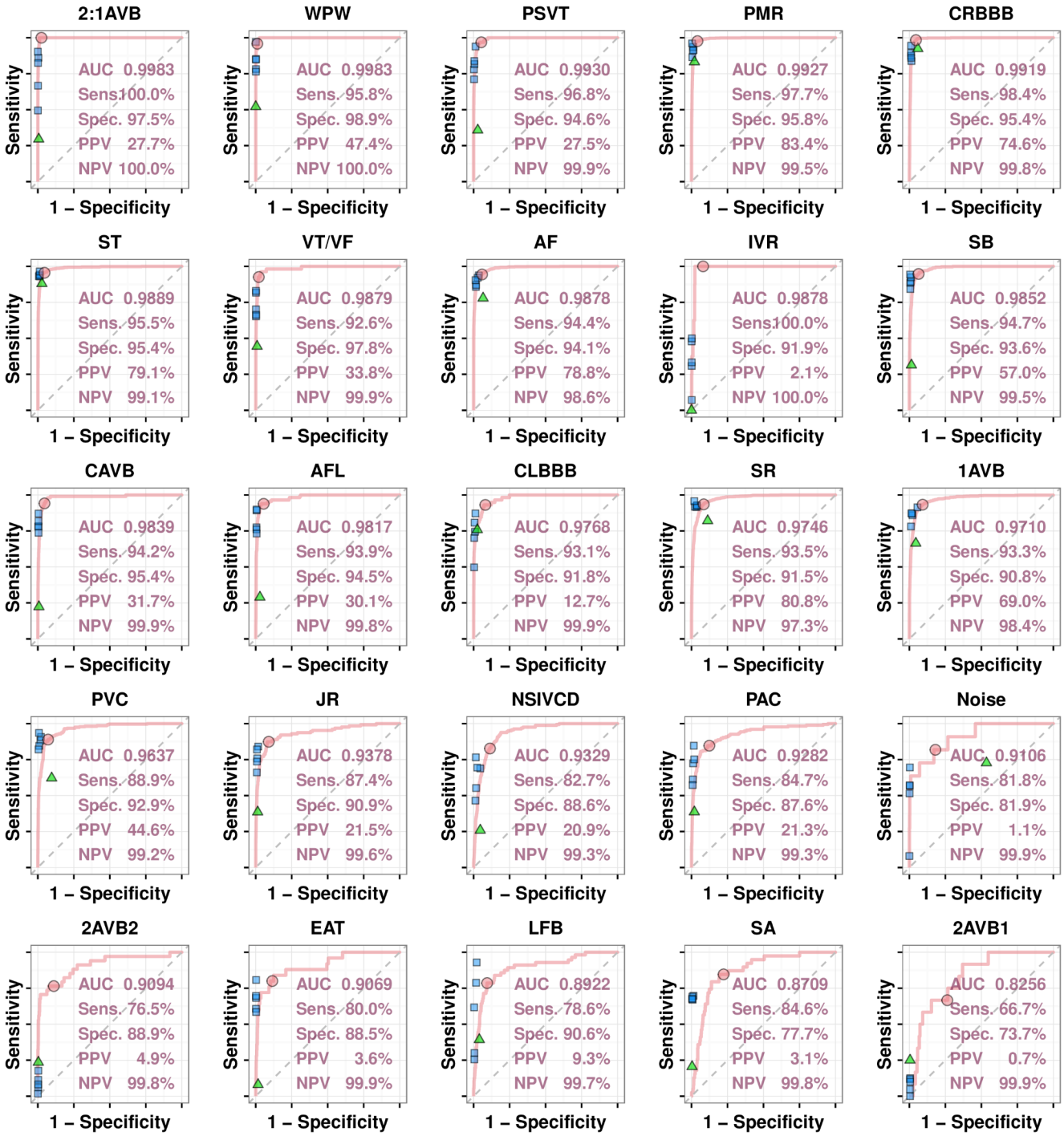


**Fig. 2** Receiver operating characteristic (ROC) curve analysis for complete 23 cardiac arrhythmias from deep learning model using 12-lead electrocardiogram (ECG) in the validation set. The operating point was selected based on the maximum of Youden's index in the tuning set and is presented using a circle mark. The area under the ROC curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated based on it. The green triangle represents the diagnosis based on the Philips ECG automatic analysis system, and the blue squares represent the performance of annotators. It should be noted that annotations from each annotator are part of the standard answer in this study

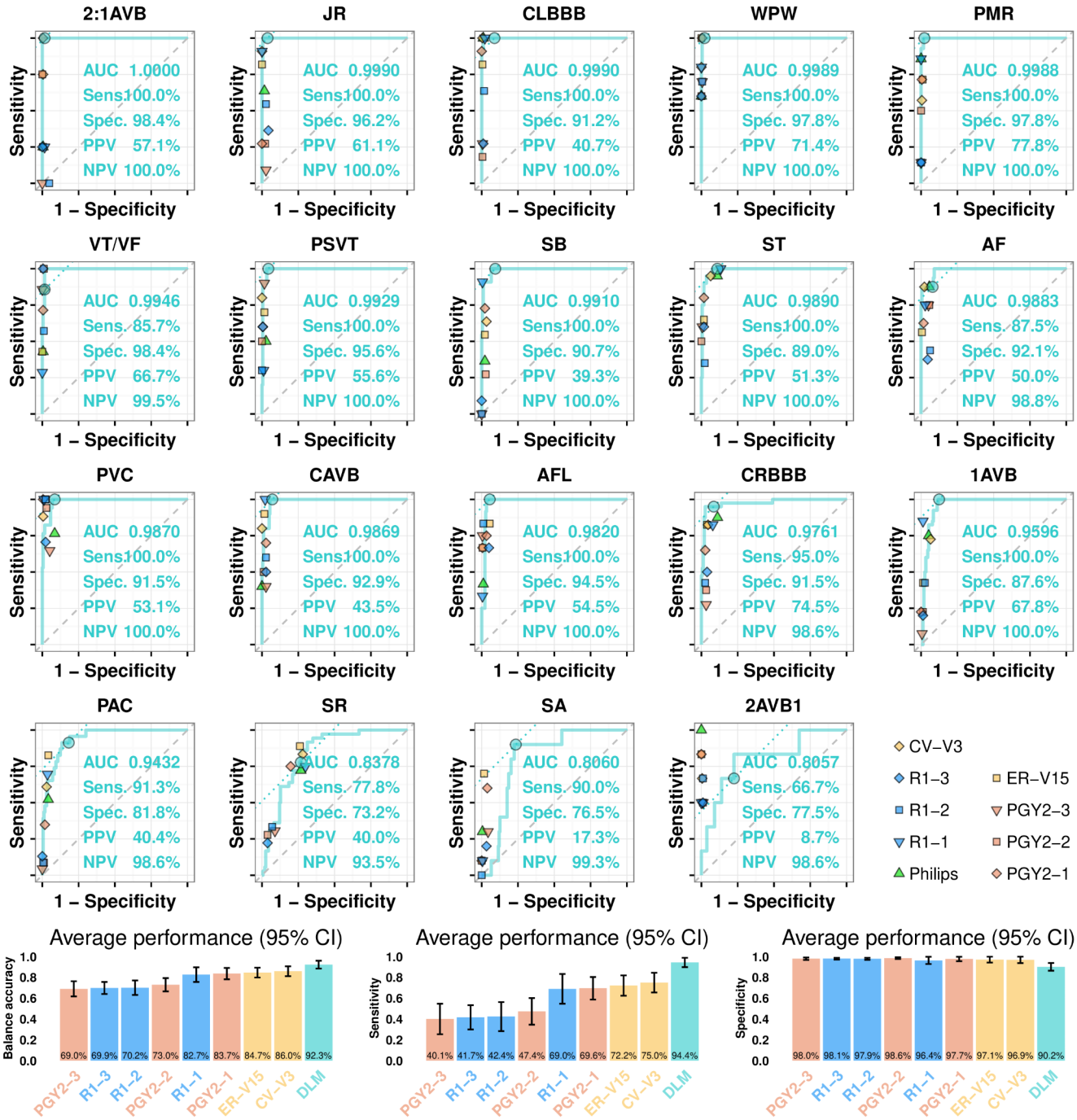


**Fig. 3** Performance comparison in recognition of 19 selected cardiac arrhythmias in the human-machine competition. All participants in this competition could access the 12-lead electrocardiogram (ECG) to decide their answers. The operating point was selected based on the maximum of Youden's index in the tuning set and is presented using a circle mark. The area under the ROC curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated based on it. The green triangle represents the diagnosis based on the Philips ECG automatic analysis system, and the other marks represent the performance of participants with different qualifications. The lower panel shows the average performance comparison in balanced accuracy, sensitivity, and specificity, ordered by balanced accuracy

resident-1 (82.7%), Philips ECG automatic analysis system (80.0%), second-year PGY-2 (73.0%), first-year resident-2 (70.2%), first-year resident-3 (69.9%), and second-year PGY-3 (69.0%). While the DLM had the best average sensitivity (94.4%) compared to other participants and the Philips ECG automatic analysis system, its average specificity was the lowest (90.2%).

Due to the subjective nature of arrhythmia interpretation and the potential for uncertainty when reading ECGs, we compared the probability distribution provided by the DLM across each annotated certainty group (Fig. 4). Group 1 comprised ECGs with a consistent arrhythmic diagnosis by all annotators; Group 2 included ECGs with initially inconsistent annotations, but whose diagnosis was subsequently confirmed by a consensus committee; Group 3 consisted of ECGs with initially inconsistent annotations, subsequently revised to another diagnosis by a consensus committee; and Group 4 included ECGs with consistent negative identification (subsequently revised to another diagnosis) from all annotators. Consequently, the certainty levels during ECG annotation were categorized as definitely positive in Group 1, probably positive in Group 2, probably negative in Group 3, and definitely negative in Group 4. The probability distribution given by the DLM consistently aligned with the annotated certainty level, from definitely positive (Group 1), probably positive (Group 2), probably negative (Group 3), to definitely negative (Group 4) in all 23 arrhythmia classes. This finding suggests a correlation between the probability distribution by the DLM and the degree of clinical uncertainty during ECG interpretation.

Figure 5 shows the lead comparison analysis for each arrhythmia class. Generally, the DLM exhibited more accurate performance by utilizing information from all 12 leads compared to using any single lead alone. Notably, lead I had a performance similar to the overall 12 leads and better performance than other single leads. The superior performance of lead I was evident in arrhythmia classes of SR, SB, ST, 1AVB, 2:1AVB, CAVB, JR, AF, WPW, NSIVCD, CLBBB, PMR, PAC, and PVC. The mean rank order was as follows: 12 leads (1.72), lead I (4.14), lead II (6.62), lead V4 (6.76), and so on.

The external validation of our DLMs in open datasets is shown in Fig. 6. In a previous study, Ribeiro et al. reported sensitivity/specificity values of 96.9%/100.0%, 93.8%/99.6%, 97.3%/99.7%, 92.9%/99.5%, 100.0%/100.0%, and 100.0%/99.5% for AF, SB, ST, 1AVB, CLBBB, and CRBBB in the CODE-test, respectively [10]. Our DLMs show comparable sensitivity/specificity of 100.0%/98.8% in AF with an AUC of 0.9996. The AUCs for SB, ST, 1AVB, CLBBB, and CRBBB were 0.9279, 0.9948, 0.9570, 0.9909, and 0.9839, respectively, compared to the corresponding AUCs in the validation set (SB = 0.9852; ST = 0.9889; 1AVB = 0.9710; CLBBB = 0.9768; CRBBB = 0.9919). For the PTB-XL dataset, the AUCs of NSIVCD, 1AVB, CAVB, CLBBB, CRBBB, and WPW were 0.8032, 0.9275, 0.9582, 0.9883, 0.9959, and 0.9651, respectively, which were consistent with another previous study, with AUCs of 0.744, 0.968, 0.967, 0.999, 0.998, and 0.855 [27]. The accuracy of these 12 arrhythmia classes in PTB-XL was similar to that in the validation set, with an AUC for AF of 0.9892, a sensitivity of 95.4%, and a specificity of 98.7%. In the CPSC2018 dataset, our DLM also achieved satisfactory AUCs of 0.9824, 0.9813, 0.9615, 0.9325, 0.8705, and 0.8865 for detecting AF, 1AVB, CLBBB, CRBBB, PAC, and PVC, respectively, which were also consistent with those in the validation set. Our DLM performed better in CLBBB compared to a previous study, with AUCs of 0.9598, [28] but the previous study reported higher AUCs of 0.9973, 0.9999, 0.9860, and 0.9647 in 1AVB, CRBBB, PAC, and PVC. The comparisons of DLMs for different types of arrhythmias across various datasets are summarized in Table 2.

To investigate the predictive ability of AI disease previvor in our DLM model, patients without histories of AF were included in the analysis. Of these, individuals who were annotated as having no AF, but labeled as having AF by the DLM were categorized into the false-positive group. Conversely, those identified as having no AF both by annotation and the DLM were categorized into the true-negative group. The long-term incidences of AF-related clinical outcomes (new-onset AF, new-onset HF, and mortality) in the two groups are depicted in Fig. 7. There were 2,650, 2,555, and 3,692 at-risk cases that developed a new-onset AF, new-onset HF, and mortality over a median (interquartile range, IQR) follow-up of 1.17 (0.34–2.53) years, 1.23 (0.52–2.89) years, and 1.43 (0.31–2.87) years, respectively. In the false-positive group, the cumulative incidence rates for new-onset AF, new-onset HF, and future mortality were 28.9%, 31.3%, and 22.2% at 2 years follow-up; 39.4%, 39.9%, and 22.2% at 4 years follow-up; and 39.4%, 39.9%, and 29.0% at 6 years follow-up. Compared with the true-negative group, patients categorized as false-positive had a significantly higher risk of developing new-onset AF (HR: 1.69, 95% CI: 1.11–2.59), new-onset HF (HR: 1.73, 95% CI: 1.20–2.51), and mortality (HR: 1.40, 95% CI: 1.02–1.92) after adjusting for sex and age.

## Discussion

Using 12-lead ECGs, we established a DLM algorithm capable of accurately detecting 23 common cardiac arrhythmias. Our study implemented a highly integrated and rigorous ECG annotation protocol and, to the best of our knowledge, included the largest number of arrhythmias compared to

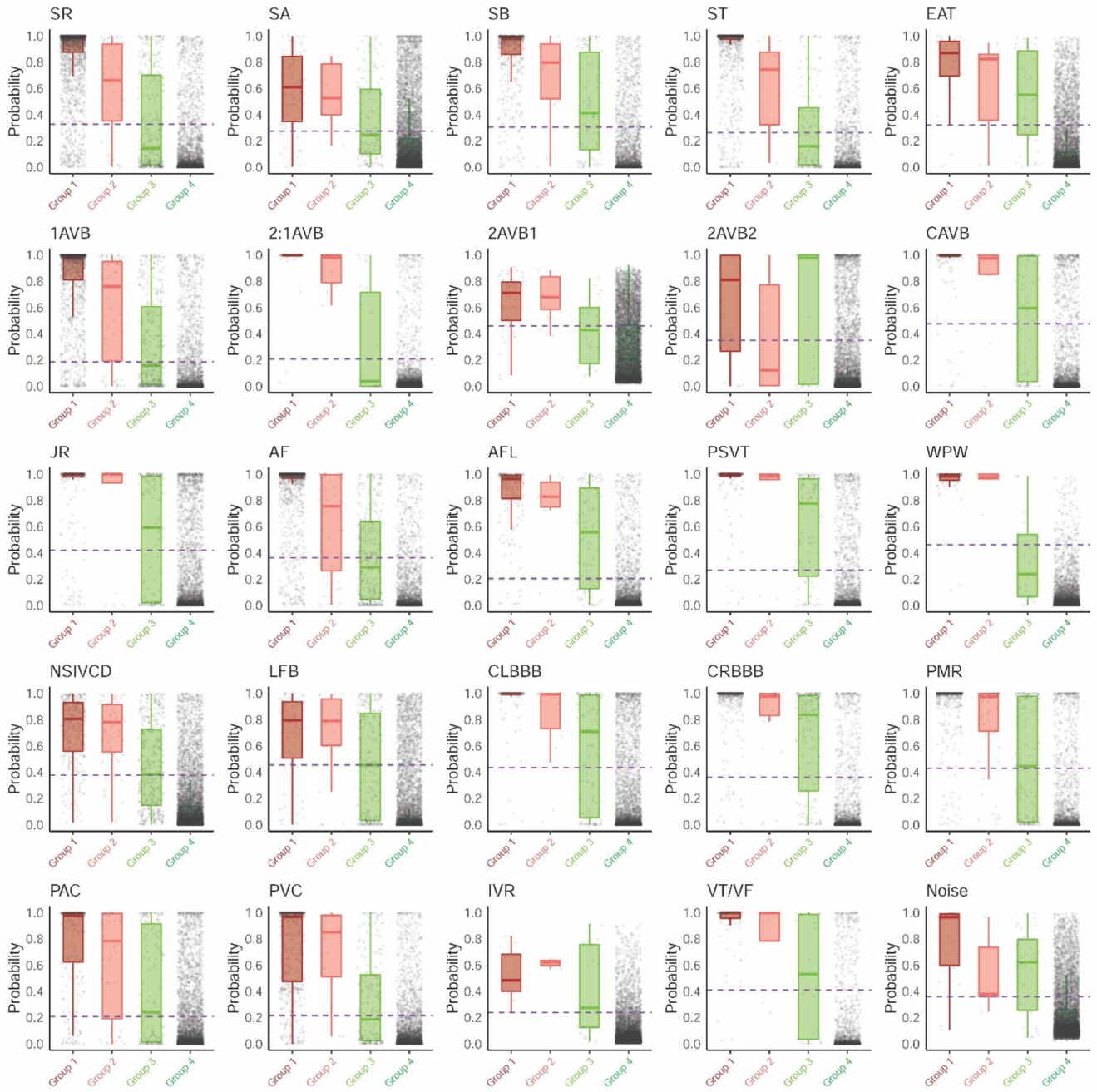


**Fig. 4** The distribution of probability given by deep learning model 12-lead electrocardiogram (ECG) in each annotated certainty group. Group 1 comprised ECGs with a consistent arrhythmic diagnosis by all annotators; Group 2 included ECGs with initially inconsistent annotations, but whose diagnosis was subsequently confirmed by a consensus committee; Group 3 consisted of ECGs with initially inconsistent annotations, subsequently revised to another diagnosis by a consensus committee; and Group 4 included ECGs with consistent negative identification (subsequently revised to another diagnosis) from all annotators. Their probabilities are presented as corresponding boxplots with jittered points. The purple dashed line is the operating point for maximizing Youden's index in the tuning set. The horizontal dotted lines represent the log-odds (or logit) of the best threshold selected from the ROC curve

previously published studies [12, 29]. Jo et al. developed an explainable deep learning model to classify nine classes of arrhythmia. Meanwhile, Zhu et al. presented a multilabel ECG diagnosis deep learning model that can detect 21 distinct arrhythmias. In contrast, our study extended its scope to encompass a broader range of potentially clinically significant arrhythmias [30]. For example, we included 2:1 AVB, a condition not previously tested in other studies. The electrocardiographic characteristics of 2:1 AVB typically manifest as a normal PR interval with a narrow QRS, followed by a

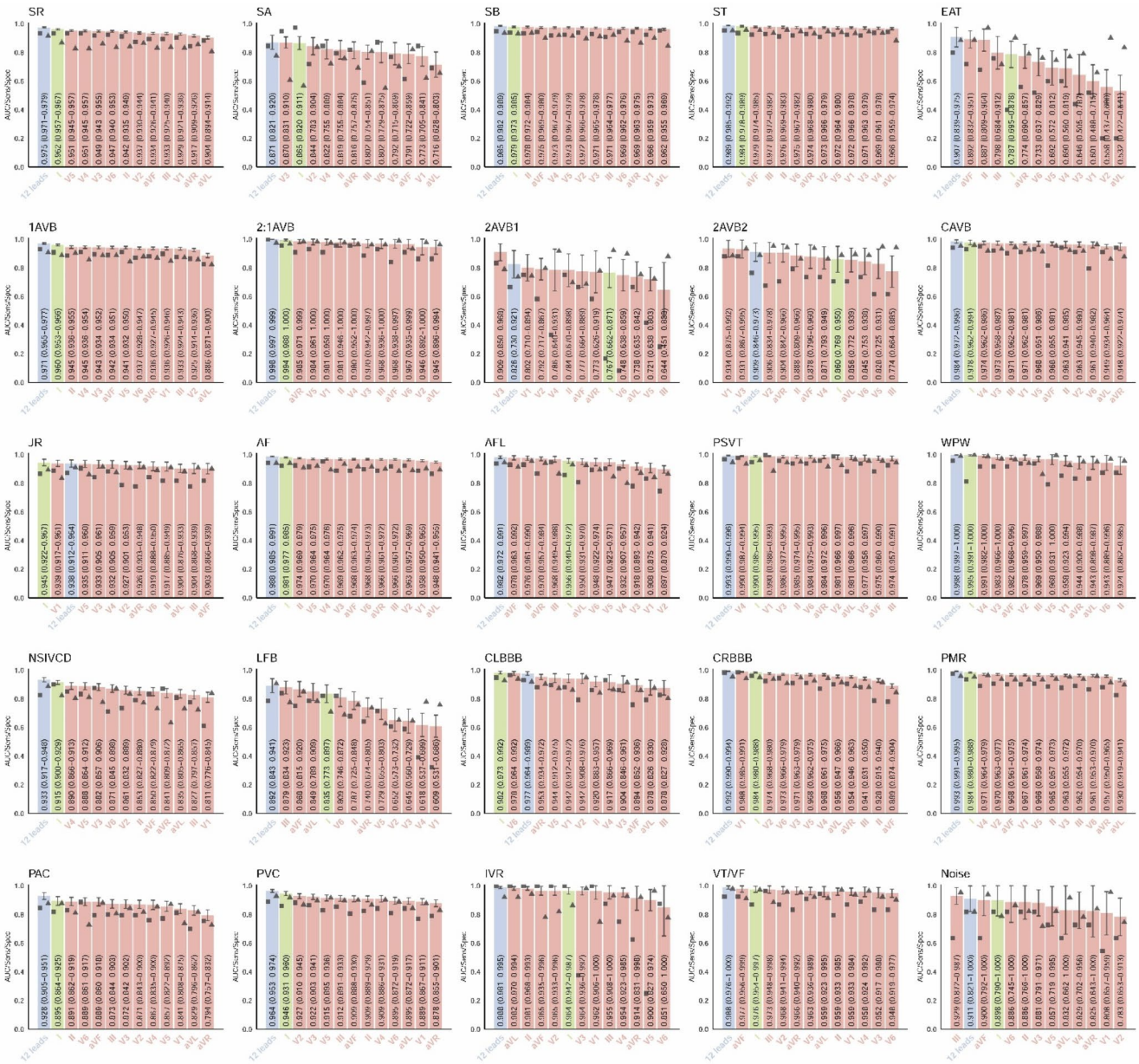


**Fig. 5** Performance ranking of deep learning model via information from 12 leads and each lead individually for detecting 23 cardiac arrhythmias in the validation set. The y-axis primarily presents the area under the receiver operating characteristic curve (AUC) as the bar and the black text with corresponding 95% confidence intervals. Sensitivity and specificity are also presented as black squares and triangles, respectively, based on the operating point for maximizing Youden's index in the tuning set. The results of 12 leads and the most popular lead, lead I, are colored as blue and yellow to emphasize their ranking

non-conducted P wave, and it requires prompt recognition and management [31, 32]. In our study, the DLM performed well in 15 arrhythmia classes with AUCs above 0.97. Specifically, the AUC for 2:1 AVB was 0.99, with a sensitivity of 100% and a specificity of 97.5%. In the human-machine competition, the DLM demonstrated cardiologist-level accuracy and achieved better performance than young residents and PGY trainees. Importantly, we also highlighted that patients with a false-positive prediction of AF exhibit a poorer prognosis than those with a true-negative prediction.

As computerized ECG analysis is routinely used worldwide, it is very convenient for clinical physicians to acquire diagnostic information from a 12-lead ECG. However, ECG misinterpretation by computerized analysis systems remains a concern. Over-reliance on computer ECG analysis can potentially result in inappropriate patient management [7]. In our study, the DLM exhibited superior diagnostic

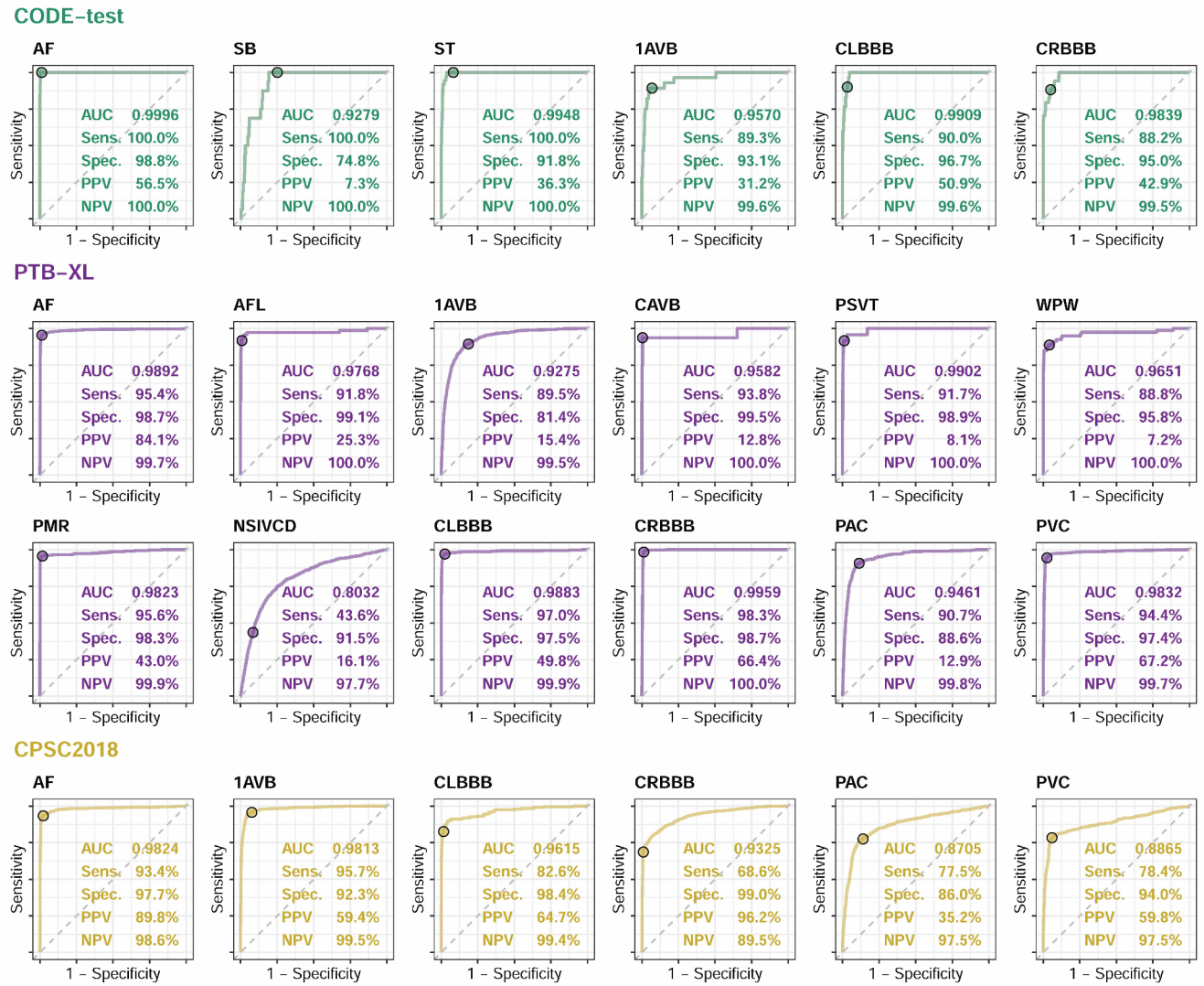


**Fig. 6** Summary of model performance in CODE-test, PTB-XL, and CPSC2018 datasets. The operating point was selected based on the maximum of Youden's index in the tuning set and is. The area under the ROC curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated based on it. This analysis was based on a complete 12-lead electrocardiogram (ECG). The labels in each dataset and our private dataset may have a definition gap or inaccurate annotations

accuracy compared to the Philips ECG analysis system in almost every class of arrhythmia. In the validation set, the DLM achieved a high AUC (>90%) for detecting 22 classes of arrhythmias. In the human-machine competition, the DLM demonstrated the best average performance in balanced accuracy compared to other competitors, including a third-year cardiologist. Notably, the DLM had the best average performance in sensitivity and balanced accuracy. The high-sensitivity operating point suggests that the DLM could be an effective screening tool to rule out a wide range of clinically significant cardiac arrhythmias, especially in high-risk patients. Due to its high negative predictive value, our DLM could assist clinical physicians by more accurately excluding normal ECGs than current computerized ECG analysis, thereby saving considerable time in interpreting large volumes of ECGs in daily practice.

Regarding the distribution of probability, we found a strong correlation between the probability given by our DLM and the degree of clinical uncertainty during ECG interpretation by physicians. In clinical practice, the interpretation of 12-lead ECGs is usually assisted by an automated computerized diagnosis, serving as an anchor for physicians. However, uncertainties are very common when there are conflicts between the interpretations made by physicians and the computerized analysis system. Notably, incorrect computerized interpretations can potentially influence the physician's thinking process and even reduce diagnostic accuracy, illustrating an "automation bias." [33].

**Table 2** Performance of models in identifying arrhythmias across open datasets

| | Model | dataset | Performance: AUC/Sens/Spec | Reference |
|---|---|---|---|---|
| AF | The present Model | CODE-test | 0.9996/100%/98.8% | |
| | Ribeiro, et al. (2020) | CODE-test | -/96.9%/100.0% | [10] |
| | The present Model | PTB-XL | 0.9892/95.4%/98.7% | |
| | The present Model | CPSC2018 | 0.9824/93.4%/97.7% | |
| | Zhou et al. (2021) | CPSC2018 | 0.9849/-/- | [28] |
| 1AVB | The present Model | CODE-test | 0.9570/89.3%/93.1% | |
| | Ribeiro, et al. (2020) | CODE-test | -/92.9%/99.5% | [10] |
| | The present Model | PTB-XL | 0.9275/89.5%/81.4% | |
| | Strodthoff, et al. (2020) | PTB-XL | 0.968/-/- | [27] |
| | The present Model | CPSC2018 | 0.9813/95.7%/92.3% | |
| | Zhou et al. (2021) | CPSC2018 | 0.9973/-/- | [28] |
| CLBBB | The present Model | CODE-test | 0.9909/90.0%/96.7% | |
| | Ribeiro, et al. (2020) | CODE-test | -/100.0%/100.0% | [10] |
| | The present Model | PTB-XL | 0.9883/97.0%/97.5% | |
| | Strodthoff, et al. (2020) | PTB-XL | 0.999/-/- | [27] |
| | The present Model | CPSC2018 | 0.9615/82.6%/84.4% | |
| | Zhou et al. (2021) | CPSC2018 | 0.9598/-/- | [28] |
| CRBBB | The present Model | CODE-test | 0.9839/88.2%/95.0% | |
| | Ribeiro, et al. (2020) | CODE-test | -/100.0%/99.5% | [10] |
| | The present Model | PTB-XL | 0.9959/98.3%/98.7% | |
| | Strodthoff, et al. (2020) | PTB-XL | 0.998/-/- | [27] |
| | The present Model | CPSC2018 | 0.9325/68.6%99.0% | |
| | Zhou et al. (2021) | CPSC2018 | 0.9999/-/- | [28] |
| PAC | The present Model | PTB-XL | 0.9461/90.7%/88.6% | |
| | The present Model | CPSC2018 | 0.8705/77.5%/86.0% | |
| | Zhou et al. (2021) | CPSC2018 | 0.9860/-/- | [28] |
| PVC | The present Model | PTB-XL | 0.9832/94.4%/97.4% | |
| | The present Model | CPSC2018 | 0.8865/78.4%/94.0% | |
| | Zhou et al. (2021) | CPSC2018 | 0.9647/-/- | [28] |

AF, atrial fibrillation; 1AVB, first-degree atrioventricular block; CLBBB, complete right bundle branch block; CRBBB, complete left bundle branch block; PAC, premature atrial complexes; PVC, premature ventricular complexes

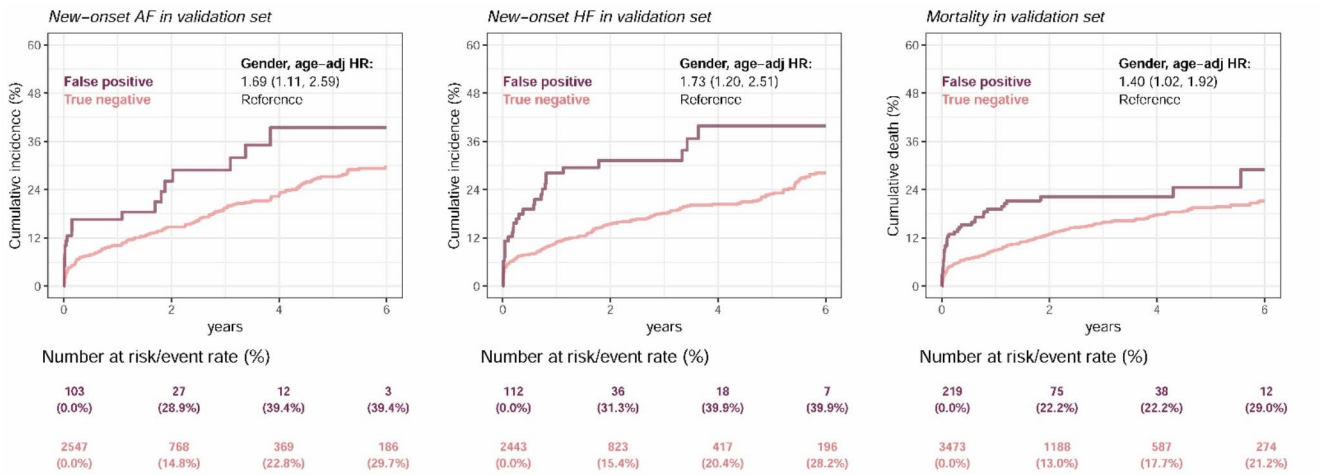


**Fig. 7** Long-term incidence of developing a new-onset disease or death in patients without atrial fibrillation (AF) initially stratified by deep learning model (DLM) classification. Long-term outcomes of patients without a corresponding disease history and present AF at the time of initial classification, stratified by the initial network classification. False positive indicates an electrocardiogram (ECG) annotated as non-AF but identified as AF by DLM. True negative indicates an electrocardiogram (ECG) annotated and DLM-identified as non-AF. The ordinate shows the cumulative incidence of developing each outcome, and the abscissa indicates years from the time of the ECG examination. The table shows the at-risk population and cumulative risk for the given time intervals in each AI stratification

Automation bias could be more pronounced in young and non-cardiology physicians, as they may have less confidence in interpreting ECGs on their own [34].

Since the probabilities output by our DLM are highly correlated with the consensus of the annotation committee for various arrhythmias, these probabilities can serve as a reference for clinical decision-making. Young and non-cardiology physicians could fully trust the AI model's suggestion if they observe probabilities > 0.9 or < 0.1. If borderline probabilities are present, it might warrant consultation with experienced experts or cardiologists. Therefore, applying our DLM-based ECG analysis in clinical practice, rather than traditional computerized ECG analysis, may aid in mitigating the impact of automation bias.

Nowadays, an increasing number of commercially available single-lead ECG devices are being used worldwide in both preclinical and clinical settings to detect and monitor cardiac arrhythmias [35]. The portable handheld single-lead ECG instruments are easy and effective single-time point screening tools for detecting arrhythmia. Users typically need to place their thumbs on two electrodes for 30 s to obtain readings. Similarly, smartwatch-based single-lead ECG devices, such as the Apple Watch, require consumers to place a finger on the Digital Crown for 30 s. Both types of devices have demonstrated comparable diagnostic accuracy to 12-lead ECGs or ambulatory 24-hour Holter monitoring in detecting atrial fibrillation [36–38]. These single-lead ECG devices have an electrical vector between one finger and the other that can simulate the classical Einthoven ECG lead I compared to standard ECGs [39]. In our lead comparison analysis, we observed that lead I performed better than other single leads in detecting a broad range of arrhythmias, including AF. Consequently, we reinforced that lead I could be a favorable choice for application in wearable single-lead ECG devices.

Apart from its application in arrhythmia detection by DLM, extracting information from an ECG beyond human recognition capabilities is also an important aspect in the era of AI. Khurshid et al. demonstrated that AI-enabled analysis of 12-lead ECGs could effectively predict the future risk of AF development [40]. Although most of our models achieved high AUCs in detecting arrhythmias, the low prevalence of certain rare arrhythmias resulted in a relatively low positive predictive value for our AI-ECG model and an elevated false-positive rate. In contrast, the false-negative rate tended to be very low, reflecting the high negative predictive value of the models. Notably, the AI-ECG's false-positive detection of atrial fibrillation may stem from frequent premature atrial complexes, potentially indicating the future development of atrial fibrillation. Clinicians should consider additional rhythm monitoring for patients with false-positive atrial fibrillation detection by AI-ECG models. In our study, individuals with false-positive AF detection by the DLM had a 1.69-fold increased risk of developing future new-onset AF, a 1.73-fold increased risk of new-onset HF, and a 1.40-fold increased risk of mortality over a 6-year follow-up period compared with true-negative individuals after adjusting for sex and age. These findings raised the concept of AI disease "previvors," referring to individuals with no ECG evidence of arrhythmia at present but with a predisposition to future risk of arrhythmic disease [25]. Our group recently found similar findings in the ECG-left ventricular diameter (ECG-LV-D) model [41]. Among individuals with echocardiography-evidenced normal left ventricular ejection fraction (LVEF), the AI-enabled ECG-LV-D model may detect subtle electrical signal changes, enabling early identification of individuals at risk of developing LV dysfunction in the future. The present AI-ECG models may serve as a valuable decision-support tool, enhancing current computerized ECG interpretations by offering more in-depth information for clinical decision-making. If frontline clinicians or cardiologists were to miss a life-threatening arrhythmia, the AI-ECG could alert them and facilitate prompt management.

## Strength and Limitations

Our study demonstrated several key strengths, including a more integrated and rigorous ECG annotation protocol than previous studies published by other research groups and a focus on a broader range of clinically significant arrhythmias, thereby expanding the usefulness of AI-ECG in clinical practice. We also conducted a human–machine competition, which clearly demonstrated the trustworthiness of our model, performed lead-by-lead comparative analyses, and externally validated our model using three open-access databases to confirm its generalizability. However, the study also had some limitations. First, this was a retrospective study in which the development sets were obtained from single-center datasets; therefore, further external validation in diverse ethnic groups and among patients with various comorbidities is warranted. Second, the study did not include certain rare but clinically significant arrhythmias, such as inherited Brugada syndrome and long-QT syndrome. Nevertheless, because the current AI-ECG models for different arrhythmias were each trained independently, new models could be seamlessly integrated in the future. Third, our DLM had higher sensitivity in most arrhythmia classes with a low positive predictive value. This indicates that clinical physicians should be alert to the possibility of arrhythmia instead of relying directly on AI-enabled diagnoses. Fourth, the potential benefit of preventive interventions for "AI disease previvors" remains unclear, necessitating further large-scale studies for validation. Finally, deep learning models still function as "black boxes," providing

limited insight into the rationale behind specific arrhythmia diagnoses, which may hinder their widespread acceptance.

## Conclusions

We developed a DLM algorithm capable of accurately detecting 23 common cardiac arrhythmias. Our DLM demonstrated superior diagnostic performance compared to the Philips ECG automatic analysis system and achieved the top position for balanced accuracy in human-machine competition. The high sensitivity of our DLM in detecting these cardiac arrhythmias suggests its potential in identifying high-risk patients, making it a promising screening tool. With interdisciplinary collaboration between clinicians and AI engineers, future research and clinical trials are warranted to evaluate the impact of AI-ECG models on clinical workflow and cardiovascular outcomes.

**Acknowledgements** We are much indebted to Sy-Jou Chen, Yi-Da Tsai, Yu-Kuang Kang, Hsin-Hui Chen, Chia-Wei Chang, Hsu-Ping Tseng, Chi-Wei Shih, and Po-Chun Lin in Tri-Service General Hospital for participating the human-machine competition.

**Author Contributions** Conception and design: Wen-Yu Lin, Chin Lin, Chih-Hung Wang, Chin-Sheng LinECG reading and annotation: Wen-Yu Lin, Wen-Cheng Liu, Chiao-Hsiang Chang, Wei-Ting Liu, Hung-Yi Chen, Chiao-Chin Lee, Yu-Cheng Chen, Chen-Shu Wu, Chin-Sheng LinCollection and assembly of data: Chin Lin, Chia-Cheng LeeData analysis and interpretation: Chin Lin, Chia-Cheng Lee, Chih-Hung Wang, Chin-Sheng Lin, Chun-Cheng LiaoManuscript writing: Wen-Yu LinFinal approval of manuscript: All authors.

**Funding** This study was supported by funding from the Ministry of Science and Technology, Taiwan (MOST110-2314-B-016-010-MY3 to C. Lin and MOST110-2321-B-016-002 to C.H. Wang and NSTC 112-2314-B-016-049, NSTC 112-2410-H-016-002 to C.S. Lin), the Tri-Service General Hospital, Taiwan (TSGH-B-111020 to C.L. Ho and TSGH-E-114215, TSGH-C01-113027 to C.S. Lin), and the Cheng Hsin General Hospital, Taiwan (CHNDMC-111-07 to C. Lin and CHNDMC-113-01, CHNDMC-113-11207 to C.S. Lin), Taichung Armed Forces General Hospital (TCAFGH-E-112052 to C.C. Liao) and National Defense Medical Center, Taiwan (MND-MAB-C02-114005 and MND-MAB-C07-113022 to C.S. Lin).

**Data Availability** No datasets were generated or analysed during the current study.

## Declarations

**Institutional Review Board** The study was conducted according to the guidelines of Declaration of Helsinki and approved by the Ethics Committee of Tri-Service General Hospital. Research ethics approval was given by the Institutional Review Board without individual consent (IRB NO. C202105049).

**Informed Consent** All data applied from quality control center were fully anonymized and encrypted from the hospital to data controller. The Institutional Review Board waived the requirement for the informed consent.

**Competing Interests** The authors declare no competing interests.

## References

1. Albert CM, Stevenson WG. The Future of Arrhythmias and Electrophysiology. Circulation. 2016;133:2687–2696. https://doi.org/10.1161/circulationaha.116.023519
2. Padeletti L, Bagliani G. General Introduction, Classification, and Electrocardiographic Diagnosis of Cardiac Arrhythmias. Card Electrophysiol Clin. 2017;9:345–363. https://doi.org/10.1016/j.ccep.2017.05.009
3. Gaztañaga L, Marchlinski FE, Betensky BP. Mechanisms of Cardiac Arrhythmias. Revista Española de Cardiología (English Edition). 2012;65:174–185. https://doi.org/10.1016/j.rec.2011.09.020
4. Al-Khatib SM, Stevenson WG, Ackerman MJ, Bryant WJ, Callans DJ, Curtis AB, Deal BJ, Dickfeld T, Field ME, Fonarow GC, et al. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death. Journal of the American College of Cardiology. 2018;72:e91-e220. https://doi.org/10.1016/j.jacc.2017.10.054
5. Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, Van Herpen G, Kors JA, Macfarlane P, Mirvis DM, et al. Recommendations for the Standardization and Interpretation of the Electrocardiogram. Journal of the American College of Cardiology. 2007;49:1109–1127. https://doi.org/10.1016/j.jacc.2007.01.024
6. Smulyan H. The Computerized ECG: Friend and Foe. The American Journal of Medicine. 2019;132:153–160. https://doi.org/10.1016/j.amjmed.2018.08.025
7. Richard H Hongo NG. Status of computerized electrocardiography. Cardiol Clin. 2006;24:491–504. https://doi.org/10.1016/j.ccl.2006.03.005
8. Teplitzky BA, McRoberts M, Ghanbari H. Deep learning for comprehensive ECG annotation. Heart Rhythm. 2020;17:881–888. https://doi.org/10.1016/j.hrthm.2020.02.015
9. Isin A, Ozdalili S. Cardiac arrhythmia detection using deep learning. Procedia Computer Science. 2017;120:268–275. https://doi.org/10.1016/j.procs.2017.11.238
10. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira W, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications. 2020;11. https://doi.org/10.1038/s41467-020-15432-4
11. Van De Leur RR, Blom LJ, Gavves E, Hof IE, Van Der Heijden JF, Clappers NC, Doevendans PA, Hassink RJ, Van Es R. Automatic Triage of 12-Lead ECGs Using Deep Convolutional Neural Networks. Journal of the American Heart Association. 2020;9. https://doi.org/10.1161/jaha.119.015138
12. Jo YY, Kwon JM, Jeon KH, Cho YH, Shin JH, Lee YJ, Jung MS, Ban JH, Kim KH, Lee SY, et al. Detection and classification of arrhythmia using an explainable deep learning model. J Electrocardiol. 2021;67:124–132. https://doi.org/10.1016/j.jelectrocard.2021.06.006
13. Mason JW, Hancock EW, Gettes LS. Recommendations for the Standardization and Interpretation of the Electrocardiogram. Circulation. 2007;115:1325–1332. https://doi.org/10.1161/circulationaha.106.180201
14. Surawicz B, Childers R, Deal BJ, Gettes LS. AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram. Journal of the American College of

Cardiology. 2009;53:976–981. https://doi.org/10.1016/j.jacc.2008.12.013
15. Ribeiro ALP, Paixão GMM, Gomes PR, Ribeiro MH, Ribeiro AH, Canazart JA, Oliveira DM, Ferreira MP, Lima EM, Moraes JLD, et al. Tele-electrocardiography and bigdata: The CODE (Clinical Outcomes in Digital Electrocardiography) study. Journal of Electrocardiology. 2019;57:S75-S78. https://doi.org/10.1016/j.jelectrocard.2019.09.008
16. Wagner P, Strodthoff N, Bousseljot R-D, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data. 2020;7. https://doi.org/10.1038/s41597-020-0495-6
17. Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, Liu Y, Ma C, Wei S, He Z, et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. Journal of Medical Imaging and Health Informatics. 2018;8:1368–1373. https://doi.org/10.1166/jmihi.2018.2442
18. KQ HGLZW. Densely connected convolutional networks. In; 2018.
19. Lin C-S, Lin C, Fang W-H, Hsu C-J, Chen S-J, Huang K-H, Lin W-S, Tsai C-S, Kuo C-C, Chau T, et al. A Deep-Learning Algorithm (ECG12Net) for Detecting Hypokalemia and Hyperkalemia by Electrocardiography: Algorithm Development. JMIR Medical Informatics. 2020;8:e15931. https://doi.org/10.2196/15931
20. Chang DW, Lin CS, Tsao TP, Lee CC, Chen JT, Tsai CS, Lin WS, Lin C. Detecting Digoxin Toxicity by Artificial Intelligence-Assisted Electrocardiography. Int J Environ Res Public Health. 2021;18. https://doi.org/10.3390/ijerph18073839
21. Lin C, Lin C-S, Lee D-J, Lee C-C, Chen S-J, Tsai S-H, Kuo F-C, Chau T, Lin S-H. Artificial Intelligence–Assisted Electrocardiography for Early Diagnosis of Thyrotoxic Periodic Paralysis. Journal of the Endocrine Society. 2021;5. https://doi.org/10.1210/jendso/bvab120
22. Liu W-T, Lin C-S, Tsao T-P, Lee C-C, Cheng C-C, Chen J-T, Tsai C-S, Lin W-S, Lin C. A Deep-Learning Algorithm-Enhanced System Integrating Electrocardiograms and Chest X-rays for Diagnosing Aortic Dissection. Canadian Journal of Cardiology. 2022;38:160–168. https://doi.org/10.1016/j.cjca.2021.09.028
23. Liu W-CL, Chin-Sheng; Tsai, Chein-Sung; Tsao, Tien-Ping; Cheng, Cheng-Chung; Liou, Jun-Ting; Lin, Wei-Shiang; Cheng, Shu-Meng; Lou, Yu-Sheng; Lee, Chia-Cheng; Lin, Chin. A deep learning algorithm for detecting acute myocardial infarction. EuroIntervention. 2021;17:765–773. https://doi.org/10.4244/EIJ-D-20-01155
24. Lin C-S, Lee Y-T, Fang W-H, Lou Y-S, Kuo F-C, Lee C-C, Lin C. Deep Learning Algorithm for Management of Diabetes Mellitus via Electrocardiogram-Based Glycated Hemoglobin (ECG-HbA1c): A Retrospective Cohort Study. Journal of Personalized Medicine. 2021;11:725. https://doi.org/10.3390/jpm11080725
25. Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. Eur Heart J. 2021;42:4717–4730. https://doi.org/10.1093/eurheartj/ehab649
26. N.V KPE. The Philips 12-Lead Algorithm Physician's Guide. https://www.documents.philips.com/doclib/enc/fetch/577817/577818/12-Lead_Algorithm_Physician_s_Guide_for_Algorithm_Version_PH080A%2C_(T-_Chinese).pdf. 2003.
27. Strodthoff N, Wagner P, Schaeffter T, Samek W. Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL. IEEE Journal of Biomedical and Health Informatics. 2021;25:1519–1528. https://doi.org/10.1109/jbhi.2020.3022989
28. Zhou H, Kan C. Tensor-Based ECG Anomaly Detection toward Cardiac Monitoring in the Internet of Health Things. Sensors. 2021;21:4173. https://doi.org/10.3390/s21124173
29. Zhu H, Cheng C, Yin H, Li X, Zuo P, Ding J, Lin F, Wang J, Zhou B, Li Y, et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. The Lancet Digital Health. 2020;2:e348-e357. https://doi.org/10.1016/s2589-7500(20)30107-2
30. Cheng S. Long-term Outcomes in Individuals With Prolonged PR Interval or First-Degree Atrioventricular Block. JAMA. 2009;301:2571. https://doi.org/10.1001/jama.2009.888
31. Barold SS, Herweg B. Second-degree atrioventricular block revisited. Herzschrittmachertherapie+Elektrophysiologie. 2012;23:296–304. https://doi.org/10.1007/s00399-012-0240-8
32. Tracy CM, Epstein AE, Darbar D, Dimarco JP, Dunbar SB, Estes NAM, Ferguson TB, Hammill SC, Karasik PE, Link MS, et al. 2012 ACCF/AHA/HRS Focused Update of the 2008 Guidelines for Device-Based Therapy of Cardiac Rhythm Abnormalities. Circulation. 2012;126:1784–1800. https://doi.org/10.1161/cir.0b013e3182618569
33. Bond RR, Novotny T, Andrsova I, Koc L, Sisakova M, Finlay D, Guldenring D, McLaughlin J, Peace A, McGilligan V, et al. Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. J Electrocardiol. 2018;51:S6-S11. https://doi.org/10.1016/j.jelectrocard.2018.08.007
34. Novotny T, Bond RR, Andrsova I, Koc L, Sisakova M, Finlay DD, Guldenring D, Spinar J, Malik M. Data analysis of diagnostic accuracies in 12-lead electrocardiogram interpretation by junior medical fellows. J Electrocardiol. 2015;48:988–994. https://doi.org/10.1016/j.jelectrocard.2015.08.023
35. Witvliet MP, Karregat EPM, Himmelreich JCL, de Jong J, Lucassen WAM, Harskamp RE. Usefulness, pitfalls and interpretation of handheld singlelead electrocardiograms. J Electrocardiol. 2021;66:33–37. https://doi.org/10.1016/j.jelectrocard.2021.02.011
36. Ramkumar S, Nerlekar N, D'Souza D, Pol DJ, Kalman JM, Marwick TH. Atrial fibrillation detection using single lead portable electrocardiographic monitoring: a systematic review and meta-analysis. BMJ Open. 2018;8:e024178. https://doi.org/10.1136/bmjopen-2018-024178
37. Rajakariar K, Koshy AN, Sajeev JK, Nair S, Roberts L, Teh AW. Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. Heart. 2020;106:665–670. https://doi.org/10.1136/heartjnl-2019-316004
38. Duarte R, Stainthorpe A, Mahon J, Greenhalgh J, Richardson M, Nevitt S, Kotas E, Boland A, Thom H, Marshall T, et al. Lead-I ECG for detecting atrial fibrillation in patients attending primary care with an irregular pulse using single-time point testing: A systematic review and economic evaluation. PLoS One. 2019;14:e0226671. https://doi.org/10.1371/journal.pone.0226671
39. Behzadi A, Sepehri Shamloo A, Mouratis K, Hindricks G, Arya A, Bollmann A. Feasibility and Reliability of SmartWatch to Obtain 3-Lead Electrocardiogram Recordings. Sensors (Basel). 2020;20. https://doi.org/10.3390/s20185074
40. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, Harrington LX, Wang X, Al-Alusi MA, Sarma G, et al. ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. Circulation. 2022;145:122–133. https://doi.org/10.1161/CIRCULATIONAHA.121.057480
41. Chen HY, Lin CS, Fang WH, Lee CC, Ho CL, Wang CH, Lin C. Artificial Intelligence-Enabled Electrocardiogram Predicted Left Ventricle Diameter as an Independent Risk Factor of Long-Term Cardiovascular Outcome in Patients With Normal Ejection Fraction. Front Med (Lausanne). 2022;9:870523. https://doi.org/10.3389/fmed.2022.870523

## Authors and Affiliations

**Wen-Yu Lin[1] · Chin Lin[2,3,4] · Wen-Cheng Liu[1] · Wei-Ting Liu[1] · Chiao-Hsiang Chang[1] · Hung-Yi Chen[5] · Chiao-Chin Lee[1] · Yu-Cheng Chen[5] · Chen-Shu Wu[5] · Chia-Cheng Lee[6,7] · Chih-Hung Wang[8,9] · Chun-Cheng Liao[10,11,12,13] · Chin-Sheng Lin[1,14]**

✉ Chun-Cheng Liao
milkbottle97@yahoo.com.tw

✉ Chin-Sheng Lin
littlelincs@gmail.com

1 Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center Taipei, Taiwan, R.O.C.

2 Medical Technology Education Center, School of Medicine, National Defense Medical Center, Taipei, Taiwan, R.O.C.

3 Graduate Institutes of Life Sciences, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

4 School of Public Health, National Defense Medical Center, Taipei, Taiwan, R.O.C.

5 Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center Taipei, Taiwan, R.O.C.

6 Medical Informatics Office, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

7 Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

8 Department of Otolaryngology-Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

9 Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan, R.O.C.

10 Department of Family Medicine, Taichung Armed Forces General Hospital, Taichung, Taiwan 411, R.O.C.

11 Department of Medical Education and Research, Taichung Armed Forces General Hospital, Taichung, Taiwan 411, R.O.C.

12 School of Medicine, National Defense Medical Center, Taipei, Taiwan 114, R.O.C.

13 No.348, Sec.2, Chungshan Rd., Taiping Dist, Taichung City, Taiwan 411228, R.O.C.

14 No 325, Section 2, Cheng-Kung Rd, Neihu, Taipei, Taiwan 11490, R.O.C.